Loss-of-Function GRHL3 Variants Detected in African Patients with Isolated Cleft Palate

Eshete, M A; Liu, H; Li, M; Adeyemo, W L; Gowans, L J J; Mossey, P A

Published in:
Journal of Dental Research

DOI:
10.1177/0022034517729819

Publication date:
2018

Document Version
Peer reviewed version

Accepted Manuscript: Eshete, M A et al. "Loss-of-Function GRHL3 Variants Detected in African Patients with Isolated Cleft Palate". Journal of Dental Research. 2018, 97(1). 41-48. https://doi.org/10.1177/0022034517729819
SAGE Publications

Loss of Function GRHL3 Variants Detected in African Patients with Isolated Cleft Palate

Mekonen A. Eshete[1, 2,3*], Huan Liu[4,5*], Mary Li[6], Wasiu L. Adeyemo[7], Lord J.J Gowans[8], Peter A. Mossey[9], Tamara Busch[6], Wakgari Deressa[1], Peter Donkor[8], Peter B. Olaitan[10], Babatunde S. Aregbesola [11], Rahman O. Braimah[12], Ganiyu O. Oseni[10], Fadekemi Oginni[11], Rosemary Audu[13], Chika Onwuamah[13], Olutayo James[7], Eno-Abasi Augustine-Akpan[6],Lukman Abdur Rahman[14], Mobolanle O. Ogunlewe [7], Fareed K.N. Arthur[8], Seidu A. Bello [15], Pius Agbenorku[8], Peter Twumasi[8], Fikre Abate[2], Taye Hailu[2], Yohannes Demissie[2, 3], Abiye Hailu [2,3], Gyikua Plange-Rhule[8], Solomon Obiri-Yeboah[8], Martine M. Dunnwald [4], Paul E. Gravem[16], Mary L. Marazita[17], Adebowale A. Adeyemo[18] , Jeffrey C. Murray[19], Robert A. Cornell[4] , Azeez Butali[6]

1) Addis Ababa University, School of Public Health, Addis Ababa, Ethiopia
2) Yekatit 12 Hospital Medical College, Addis Ababa, Ethiopia
3) Department of Surgery, Addis Ababa University School of Medicine, Ethiopia
4) Department of Anatomy and Cell Biology, University of Iowa, Iowa City. USA
5) State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) and Key Laboratory for Oral Biomedicine of Ministry of Education, School and Hospital of Stomatology, Wuhan University, Wuhan, China
6) Department of Oral Pathology, Radiology and Medicine, University of Iowa, USA
7) Department of Oral and Maxillofacial Surgery, University of Lagos, Lagos, Nigeria
8) Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
9) Department of Orthodontics, University of Dundee, Dundee, UK
10) Department of Plastic Surgery, Ladoke Akintola University of Science and Technology, Osogbo, Nigeria
11) Department of Oral and Maxillofacial Surgery, Obafemi Awolowo University, Ile Ife, Nigeria
12) Department of Oral and Maxillofacial Surgery, Usmanu Danfodiyo University, Sokoto, Sokoto, Nigeria
13) Department of Virology, Nigerian Institute of Medical Research, Lagos, Nigeria
14) Division of Pediatric Surgery,Department of Surgery, University of Ilorin, Ilorin, Nigeria
15) State House Clinic, Abuja, Nigeria
16) Haukeland University Hospital, Plastic and Reconstructive Surgery Department, Bergen, Norway.
17) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, U.S.A
18) National Human Genomic Research Institute, Bethesda, Maryland, USA
19) Department of Pediatrics University of Iowa, Iowa City, USA

[*]Both authors contributed equally to the study

Authors for correspondence:

A) Dr. Azeez Butali, Department of Oral Pathology, Radiology and Medicine, College of Dentistry, University of Iowa, Iowa City IA 52241. Email: Azeez-butali@uiowa.edu, Fax: 319-384-1169. Tel: 319-335-8980

B) Dr. Robert Cornell, Email: robert-cornell@uiowa.edu, Tel: 319-335-8908

Abstract

In contrast to the progress that has been made towards understanding the genetic etiology of cleft lip with or without cleft palate (CL/P), relatively little is known about the genetic etiology for cleft palate only (CPO). Recently**,** a common coding variant of Grainyhead Like Transcription Factor 3 (GRHL3) was shown to be associated with risk for CPO in Europeans. Mutations in this gene had also been previously reported in families with Van der Woude syndrome (VWS). To identify rare mutations in *GRHL3* that might explain the missing heritability for CPO, we sequenced *GRHL3* in cases with CPO from Africa. We recruited participants from Ghana, Ethiopia and Nigeria. This cohort included case-parent trios, cases and other family members, as well as controls. In total we sequenced exons of this gene in DNA from 134 non-syndromic cases. When possible, we sequenced them in parents to identify de novo mutations. Five novel mutations were identified including two missense (c.497C>A; p.Pro166His and c.1229A>G; p.Asp410Gly), a splice site (c.1282A>C p.Ser428Arg), a frameshift (c.470delC; p.Gly158Alafster55), and a nonsense mutation (c.1677C>A; p.Tyr559Ter)). These mutations were absent from 270 sequenced controls and from all public exome and whole genome databases, including the 1000 genomes database (which includes data from Africa). However, four of the five mutations were present in unaffected mothers, indicating that their penetrance is incomplete. Interestingly, one mutation damaged a predicted sumoylation site, and another disrupted a predicted CK1 phosphorylation site. Over-expression assays in zebrafish and reporter assays in vitro indicated that four variants were functionally null or hypomorphic while one was dominant-negative. This study provides evidence that, as in Caucasian populations, mutations in *GRHL3* contribute to the risk of non-syndromic CPO in the African population.

Key words: sub-Saharan Africa, targeted sequencing, GWAS, zebrafish, missense mutation

Introduction

Orofacial clefts (OFC) are common in humans and are the most common craniofacial birth defects, with a worldwide prevalence of 1 in 700 live births (Mossey and Little 2002). They are usually divided into syndromic and non-syndromic clefts. Non-syndromic clefts, which account for about 70% of all clefts, occur without additional visible clinical features. Syndromic clefts are associated with additional structural abnormalities. The cleft spectrum that includes cleft palate only (CPO), cleft lip only (CLO) and cleft lip with or without palate (CL/P) can be found in either syndromic or non-syndromic clefts. About 50% of CPO is classified as syndromic clefts (Marazita et al. 2002). Regardless of the category or types of clefts, the rehabilitation of affected individuals requires a multidisciplinary team approach. For instance, in the U.S., it has been estimated that the lifetime cost of managing a child with cleft may total as much as $100,000 (Waitzman et al., 1994). Therefore, OFCs places a huge personal and financial burden on families, society and healthcare systems, particularly in developing countries where social and physical infrastructures are limited.

The etiology of non-syndromic cleft palate is complex and factors including genetics, environmental exposures and gene-environment interaction are plausible causes (Beaty et al., 2016). Unlike nonsyndromic CL/P where over 26 risk loci and several candidate genes have been implicated through some genome-wide association studies (GWAS) (Beaty et al., 2010; Ludwig et al., 2012; Sun et al., 2015; Leslie et al., 2016a; Leslie et al.,2017; Yu et al.,2017), little is known about the genetic causes of CPO. Ghassibe et al in 2011 reported a role for FAS-associated factor 1 (FAF1) in CPO when they demonstrated that haploinsufficency of FAF1 causes CPO in a family with Pierre Robin Sequence (PRS) (Ghassibe et al. 2011). Pierre Robin Sequence is characterized by CPO, micrognathia and glossoptosis. In another study, Kilpatrick et al. in 2003 reported two de novo CPO-associated translocations involving 2q32-q33 where one breakpoint obstructs the transcription unit of Special AT-rich sequence-binding protein 2

(SATB2) (Kilpatrick et al. 2003). Also, mutations in T-Box 22 (TBX22) have been reported in X-linked CPO cases (Braybrook et al. 2001; Braybrook et al. 2002; Marcano et al. 2004).

Two recent independent studies identified a common missense mutation, p.Thr454Met, in *GRHL3* that is associated with CPO (Leslie et al. 2016b; Mangold et al. 2016). The first was a case-control/case-triad GWAS (Leslie et al., 2016b); and the second was a combination of targeted sequencing and association study (Mangold et al. 2016). Mutations in GRHL3 have been previously reported in VWS cases with CPO (Peyrard-Janvid et al. 2014). *GRHL3* encodes the transcription factor Grainyhead-like 3, which is necessary for formation of the epidermal permeability barrier in mice (Ting et al. 2005). Both it and the transcription factor Interferon Regulatory Factor 6 (IRF6) are necessary for differentiation of oral periderm, the most superficial layer of oral epithelium that covers palate shelves during morphogenesis of the face (Peyrard-Janvid et al. 2014; Ingraham et al. 2006). Loss of oral periderm results in adhesions between palate shelves in mice (Richardson et al. 2009); leading to the proposal that defective differentiation of oral periderm is the cellular event that ultimately results in cleft palate. In zebrafish, two homologues of *GRHL3*; *Grhl1* and *Grhl3*, acting redundantly and the ortholog of *IRF6* are required for differentiation of periderm, the most superficial layer of embryonic skin, with the Grhl factors acting as a downstream of Irf6 (Sabel et al. 2009; de la Garza et al. 2013). Injection of RNA encoding human GRHL3 mRNA into zebrafish eggs induced ectopic expression of a keratin 4 (*krt4*), a marker of the periderm, in deep blastomeres at gastrula stage (de la Garza et al. 2013); this assay was used to show that VWS-associated and non-syndromic CPO-associated variants of GRHL3 lack normal function (Peyard et al. 2014; Leslie et al., 2016b). In both cases, disease-associated variants disrupted differentiation of periderm in wild-type embryos, indicating they had dominant-negative activity.

Importantly, the common variant p.Thr454Met in GRHL3 explains only a fraction of the heritability for CPO. To identify rare variants that could explain in part the heritability for CPO, we conducted targeted sequencing of GRHL3 gene in individuals with CPO from Ghana, Ethiopia and Nigeria. We tested novel variants found in patients in functional assays in cell lines and in zebrafish embryos.

Subjects and Methods

This study is part of a collaborative study that investigates the genetic and environmental causes of orofacial clefts in sub-Saharan Africans from Ghana, Ethiopia and Nigeria. Local Institutional Review Boards (IRBs) approved sample and data collection: Ethiopia-Institutional Review Board College of Health Sciences, Addis Ababa University IRB No 3.10/027/2015; Kwame Nkrumah University of Science and Technology, Kumasi, Ghana– IRB No CHRPE/AP/217/13; and College of Medicine University of Lagos, Nigeria – IRB No ADM/DCST/HREC/APP/1374. Patients born with cleft palate only, controls and parents of both cases and controls were recruited and assessed at treatment centers in each country by the investigators and their collaborators. Informed consent was obtained from all participating families and saliva samples were collected using Oragene saliva kits and sponges (DNA Genotek, Kanata, ON, Canada). These samples were sent to the Butali Laboratory at the University of Iowa, USA, for processing and analysis. We extracted DNA from all samples followed by XY genotyping analysis to ensure the gender of the sample matched the gender of the donor, and for quality control purposes. A combination of 134 case-parent samples (dyads and triads), as well as 270 unrelated controls (90 from each of the three countries), were included for Sanger sequencing.

Sanger Sequencing

For sequencing DNA from human subjects we used methods that we reported previously (Gowans et al. 2016). We used primers that were optimized for the amplification of 17 exons in the GRHL3 gene (NM_198174). We used 4 ng of DNA in a 10 ul reaction for the Polymerase chain reaction (PCR). Two Yoruba HapMap samples and two water samples were added to the 96 well plates as template and non-template controls, respectively. The primers used and annealing temperatures are available from the Butali Laboratory upon request. The amplified DNA products were sequenced at Functional Biosciences (http://order.functionalbio.com/seq/index) (Madison, WI).

To recognize novel mutations, we compared mutations that we identified in patients with those in the 1000 genomes (1KG) database (http://www.1000genomes.org/), Exome variant sequence (EVS) database (http://snp.gs.washington.edu/EVS/), and Exome Aggregate Consortium (ExAC) database (http://exac.broadinstitute.org/). These databases included exome data for over 5200 African and African American controls. We also sequenced 270 controls as an additional step to confirm novel mutations. To predict the functional effects of novel mutations on the protein, we used bioinformatics tools such as Polymorphism Phenotyping (Polyphen) (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al. 2010), Sorting Intolerant From Tolerant (SIFT) (http://sift.jcvi.org/) (Kumar et al. 2009), and Have Your Protein Explained (HOPE) (http://www.cmbi.ru.nl/hope) (Venselaar et al. 2010). To assess inheritance of novel mutations, we sequenced parent samples, when available.

Functional test of human GRHL3 mutations in zebrafish embryos

Full-length, wild type human *GRHL3* cDNA (GenBank: BC036890.2) was obtained from GE Healthcare (Pittsburgh, PA). Novel *GRHL3* mutations were introduced into the wild type using PCR-mediated mutagenesis and the cDNAs shuttled into CS2+ vector (gift of David Turner, University of Michigan). The corresponding capped mRNAs were generated *in vitro* with the mMESSAGE mMACHINE SP6 kit (Ambion, Grand Island, NY) and purified with the RNeasy mini kit (Qiagen, Valencia, CA). For "rescue" experiment, 5 nl of *grhl1* AUG MO and/or of *grhl3* E4I4 MO, each at 1 mg/ml, were injected into wild-type zebrafish embryos (NHGRI line) at the single cell stage after which approximately 1 ng different human *GRHL3* mRNAs were injected into the same embryos. Embryos were fixed at 8 hours post fertilization (hpf) and whole-mount in situ hybridization for *krt4* was performed as described previously (Sabel et al., 2009). For cryo-section, after in situ hybridization embryos were embedded in Tissue-Tek® O.C.T. Compound (Sakura Finetek, USA) at -80°C. 10 micron sections were made using Microm Cryostat I HM505E (GMI, Ramsey, MN). The Public Health Service Assurance approved animal use protocols (protocol No. 6011616). Details of RNA extraction and luciferase assay are included in Appendix.

Results

Novel *GRHL3* mutations detected in patients with cleft palate only

We sequenced *GRHL3* exons and splice sites in 134 cases with non-syndromic CPO and none in 270 controls. We identified five novel variants in patients that were absent from controls, from the ExAc database of more than 100,000 whole exomes, and the 1000 Genomes

database of whole genome sequences (accessed on 08/04/2017). The mutations included 2 missense (p.Pro166His and p.Asp410Gly, referring to *GRHL3* isoform 2, NP_937816), one splice site (p.Ser428Arg), a deletion (p.Gly158Alafster55), and one nonsense mutation (p.Tyr559Ter) (Table 1, Fig. 1A). All of the identified mutations except for p.Ser428Arg were also detected in an unaffected parent. For the p.Ser428Arg mutation, we only have samples for the mother who did not have the mutation (Table 2). We also identified known missense mutations in cases (Appendix Table 1). The five novel variants have been deposited into the Leiden Open Variation Database with submission number 00100262.

CPO patient-derived variants of GRHL3 have lower-than-normal ability to induce ectopic expression of a periderm marker in zebrafish embryos.

We first tested the activity of CPO-patient derived *GRHL3* variants in an over-expression paradigm. We engineered each of variants into full length *GRHL3* cDNA, synthesized mRNAs in vitro, and injected the mRNAs (or lacZ mRNA as a negative control) singly into single-cell zebrafish embryos, fixed embryos at 6 hours post-fertilization (hpf)(shield stage), and processed them to reveal *krt4* expression by in situ hybridization (Peyard-Janvid et al. 2014; Leslie et al. 2016b). All embryos injected *lacZ* mRNA exhibited *krt4* expression solely in the enveloping layer (EVL), which is the normal pattern of expression (40 of 40 embryos injected) (Rauch et al. 2003) (Fig. 1B; Appendix Fig.1). In contrast, all embryos injected with wild-type human *GRHL3* mRNA exhibited large patches of ectopic expression of *krt4* in deep blastomeres readily visible in a whole-mount preparation (Fig. 1C) and confirmed by in sectioned embryos (Appendix Fig. 1) (40 of 40 injected embryos). In embryos injected with the p.Gly158Alafster55 (Fig. 1D), p.Asp410Gly (Fig. 1F) and p.Ser428Arg (Fig. 1G) variants there were no ectopic patches of *krt4* expression (more than 60 embryos for each construct). In contrast, a small fraction (<10%) of

embryos injected with the p.Pro166His (Fig. 1E) and p.Tyr559Ter (Fig. 1H) variants exhibited small patches of ectopic *krt4* expression (4 of 88, and 5 of 91 embryos injected, respectively), revealing that in this assay these variants are hypomorphic (i.e., less active than wild-type). Moreover, in about half of embryos injected with the p.Ser428Arg variant, there were clear gaps in the expression of *krt4* (23 of 64 embryos injected and processed for *krt4* expression) (Fig. 1G); one of the embryos injected with this construct ruptured just prior to fixation (i.e., at 8 hpf). Therefore we repeated the experiment with this variant and permitted embryos to continue to develop. By 9 hpf, 36 of 83 embryos injected had ruptured similar to embryos injected with a dominant-negative variant of *Grhl3* (de la Garza et al. 2009). In summary, the results indicate that in the over expression assay, the p.Pro166His and p.Tyr559Ter variants are hypomorphic, p.Gly158Alafster55 and p.Asp410Gly variants are strongly hypomorphic or null, and the p.Ser428Arg variant is dominant-negative.

Next, in a rescue paradigm, we asked whether *krt4* expression could be restored by injection of mRNA encoding GRHL3 variants. Previously we showed that simultaneous inhibition of *grhl1* and *grhl3* using morpholinos disrupts epiboly and *krt4* expression, but knockdown either of gene singly does are not grossly affect periderm development (de la Garza, et al. 2009) (Fig. 2A-D). For rescue assays, we injected human *GRHL3* variants into embryos depleted of both *grhl1* and *grhl3*. Injection of wild-type *GRHL3* into embryos injected with *grhl1* MO and *grhl3* MO partially restored superficial *krt4* expression and induced ectopic expression of *krt4* in deep blastomeres (Fig.2 E). For quantitative comparisons of rescue efficiency, we evaluated *krt4* expression in whole embryo lysates by qRT-PCR (Fig. 2F). Embryos injected with *grhl1* MO and *grhl3* MO (i.e., double morphants) injected with *LACZ* mRNA had significantly lower *krt4* expression than those injected control morpholino, whereas double-morphants injected with RNA coding the

reference (wild type) variant, the p.Pro166His variant, or the p.Tyr559Ter variant of GRHL3 had significantly higher levels of *krt4* expression than those injected with *LACZ* mRNA. *krt4* expression levels in double morphants injected with the other variants (p.Gly158Alafster55, p.Asp410Gly and p.Ser428Arg) were not significantly higher than those in embryos injected with LACZ, and were significantly lower than in those injected with wild-type GRHL3 ($p<0.01$ by one-way ANOVA) (Fig. 2F).

CPO patient-derived variants have lower-than-normal ability to activate a GRHL3 sensitive reporter

We next assessed the ability of the GRHL3 variants to activate a synthetic GRHL3-sensitive reporter, which was comprised of 6 GRHL3 binding sites upstream of a minimal promoter and the firefly luciferase gene (Leslie et al. 2016b). We co-transfected 293FT cells with the GRHL3-sensitive reporter, expression vectors harboring different GRHL3 variants (separately), and a plasmid driving constitutive Renilla luciferase expression as a transfection control (internal control). As controls, we additionally tested other patient-derived GRHL3 variants, both previously shown to have dominant-negative activity in zebrafish: p.Arg391Cys, from an individual with VWS (Peyard-Janvid et al. 2014) and p.Thr454Met, a common variant associated with risk for nsCPO (Leslie et al., 2016b Mangold et al., 2016). Compared to cells transfected with the GRHL3-sensitive reporter alone, those additionally transfected with the expression vector containing wild-type GRHL3 had more than 6-fold greater luciferase levels (Fig. 3). By contrast, expression vectors containing the p.Ser428Arg did not activate the reporter above background, and those containing p.Pro166His, p.Tyr559Ter and p.Thr454Met variants did so but to a significantly lesser level than did wild-type (Fig. 3A). Of note, the

p.Pro166His and p.Tyr559Ter were the variants that also retained some ability to induce ectopic *krt4* expression in zebrafish embryos (Fig. 1E,H).

p.Ser428Arg GRHL3 variant shows dominant negative effects in 293FT cells.

Finally, to test the patient-derived GRHL3 variants for dominant-negative activity, we co-transfected equal amounts of expression vectors containing wild-type and a patient-derived variant. The p.Ser428Arg, p.Arg345Cys and p.Thr408Met, which all had dominant-negative activity in the zebrafish assay, also all decreased the reporter levels compared to cells transfected with wild-type GRHL3 alone (Fig. 3B). The remaining variants all had additive effects when combined with wild-type GRHL3, indicating they are loss-of-function but not dominant-negative variants. In summary, variants of GRHL3 detected in patients with CPO all had reduced function in comparison to the reference variant, and a subset had dominant-negative activity, consistent with their having pathogenic effects.

Discussion

Based on evidence on the association of a variant of *GRHL3* with CPO in Europeans, we tested the hypothesis that rare-function blocking variants of *GRHL3* contribute to pathogenesis of NS CPO in Sub Saharan Africa by sequencing the GRHL3 exons in such patients. We detected 5 variants that were of interest because of their absence from public domain catalogs of coding variants and because one or more in silico algorithm predicted they were “possibly damaging” or worse. We subjected all of them to functional assays in zebrafish, including an over-expression and a rescue of *grhl3* loss-of-function paradigm, and to in vitro tests of their transactivation activity, all in comparison to the reference variant of GRHL3. These

assays consistently indicated that two variants, p.Pro166H and p.Tyr559Ter, were modestly hypomorphic, two others, p.Gly158Alafster55 and p.Asp410Gly, were strongly hypomorphic or null, and one, p.Ser428Arg, was dominant negative. These findings raise several issues as follows.

First, the presence of some of these mutations in unaffected family members suggests that their effects are completely penetrant, as has been reported in previous genetic studies of clefting (Mangold et al. 2016). Modifier alleles in the parents may be protective for clefting but simultaneously elevate their risk for other chronic diseases like cancer which has elevated prevalence in families with clefts (Christensen et al. 2004; Bille et al. 2005). Alternatively, affected children and unaffected parents may share the same genetic risk factors but may have been differentially exposed to environmental factors, such as maternal smoking during the periconceptional period, leading to clefts in the presence of risks alleles (Little et al. 2004).

Second, it is noteworthy that our functional assays had only modest concordance with *in silico* algorithms that predict the functional consequence of missense variants, e.g., Polyphen, SIFT, and HOPE. The most extreme disagreement was for the p.Ser428Arg variant, which had dominant-negative activity in our assays, but was predicted to be benign by both Polyphen and SIFT. The p. Pro166His mutation, which had hypomorphic activity in our in vivo and in vitro assays, was predicted to be “probably damaging” by Polyphen and “tolerated” by SIFT with a Provean score of -0.64. The p.Tyr559Ter variant, which had reduced transactivation activity in our in vitro assay, was predicted to be a gain of function mutation by HOPE leading to increased activity of GRHL3. The p.Asp410Gly variant, which in our assays was loss-of-function, Polyphen predicted to be probably damaging and SIFT predicted to be deleterious. The strength

of the zebrafish-based assays is that they tested the variants function in the context of a gene regulatory network governing differentiation of a relevant tissue. The in vitro assays have the advantage of being more quantitative than the in vivo ones. The results from in vivo assay is consistent with the two in vitro assays, which agreed with one another, all supporting their validity. Functional assays remain an important complement to in silico algorithms for discerning pathological variants from innocuous ones.

Third, the specific biochemical deficits in each functional variant are unknown and are an interesting subject for future study. One or more of the variants may diminish the protein's stability which is testable with appropriate antibodies or epitope-tagged variants. HOPE predicted that p.Ser428Arg disrupts a potential phosphorylation site of the kinase CK1, although it is unknown if GRHL3 is phosphorylated at this residue. The same algorithm predicted that the substitution of proline by histidine in the p.Pro166His variant will lead to a loss of hydrophobic interaction in the nucleus or surface of the protein, and prolines provide rigidity to local protein conformation. Finally, it predicted that p.Asp410Gly disrupts a potential sumoylation site; sumoylation of lysine residues generally alters protein stability, function, or subcellular localization (Hendrik and Vertegaal 2015). It remains to be determined if this residue of GRHL3 is subjected to sumoylation.

Conclusion

We identified novel GRHL3 variants in CPO patients and tested them in functional assays that had the ability to recognize wild-type, reduction in function, and dominant-negative activities. The results of the functional tests suggest that lower-than-normal levels of GRHL3

and not necessarily dominant negative, activity can confer increased susceptibility to CPO. The p.Thr454Met variant, first identified in a European GWAS was not found in our samples, but other rare variants of GRHL3 influence heritability for CPO in Africans. These findings expand understanding of the genetics and biology underpinning CPO.

Acknowledgements

We appreciate the support of all members of the family who participated in this research. This work was supported by NIDCR K99/R00 Grant DE022378 and Robert Wood Johnson Foundation Grant number 72429 (AB), NIH R01 DE023575 (RAC), R37 grants DE-08559 and DE-016148 (JCM), NIH AR067739 R01 (MD), and Ghana Cleft Foundation, G001 (LJJG).

We have no conflicts of interest to declare.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense variants. Nat Methods. 7(4)**:**248-249.

Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, et al. 2010. A genome wide association study of cleft lip with / without cleft palate using case-parent trios of European and Asian ancestry identifies *MAFB* and *ABCA4* as novel candidate genes. Nat Genet. 42 (6):525-529.

Beaty TH, Marazita ML, Leslie EJ. 2016. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. F1000Research.5:2800.

Bille C, Winther JF., Bautz A, Murray JC., Olsen J, Christensen K. 2005. Cancer Risk in Persons with Oral Cleft - A Population-based Study of 8,093 Cases. Am J Epidemiol. 161(11):1047-1055.

Braybrook C, Doudney K, Marcano ACB, Arnason A, Bjornsson A,Patton MA, Goodfellow PJ, Moore GE, Stanier P. 2001. The T-box transcription factor gene TBX22 is mutated in X-linked cleft palate and ankyloglossia. Nat Genet. 29 (2):179–183.

Braybrook C, Lisgo S, Doudney K, Henderson D, Marc¸ano ACB, Strachan T,Patton MA, Villard L, Moore GE, Stanier P, et al. 2002. Craniofacial expression of human and murine TBX22 correlates with the cleft palate and ankyloglossia phenotype observed in CPX patients. Hum Mol Genet. 11(22):2793–2804.

Butali A, Mossey PA, Adeyemo WL, Eshete MA, Gaines LA, Even D, Braimah RO, Aregbesola SB, Rigdon JV, Emeka CI et al. 2014. Novel IRF6 Mutations in Families with Van Der Woude Syndrome and Popliteal Pterygium Syndrome from Sub-Saharan Africa. Mol Genet Genomic Med 2(3):254-260.

Christensen K, Juel K, Herskind AM, Murray JC. 2004. Long term follow up study of survival associated with cleft lip and palate at birth. BMJ. 328(7453):1405.

de la Garza G, Schleiffarth JR, Dunnwald M, Mankad A, Weirather JL, Bonde G, Butcher S, Mansour TA, Kousa YA, Fukazawa CF, et al. 2013. Interferon regulatory factor 6 promotes differentiation of the periderm by activating expression of Grainyhead-like 3. J Invest Dermatol.133 (1):68-77.

FitzPatrick DR, Carr IM, McLaren L, Leek JP, Wightman P, Williamson K, Gautier P, McGill N, Hayward C, Firth H, et al. 2003. Identification of SATB2 as the cleft palate gene on 2q32-q33. Hum Mol Genet. 12 (19): 2491-2501.

Ghassibe-Sabbagh M, Desmyter L, Langenberg T, Claes F, Boute O, Bayet B, Pellerin P, Hermans K., Backx L, Mansilla MA., et al. 2011. FAF1, a gene that is disrupted in cleft palate and has conserved function in zebrafish. Am. J. Hum.Genet. 88(2), 150–161.

Gowans LJJ, Adeyemo WL, Eshete M, Mossey PA, Busch T, Aregbesola B, Donkor P, Arthur FK, Bello SA, Martinez A, et al. 2016. Association Studies and Direct DNA Sequencing

Implicate Genetic Susceptibility Loci in the Etiology of Nonsyndromic Orofacial Clefts in Sub-Saharan African Populations. *J Dent Res*. 95(11):1245-1256.

Hendriks IA, Vertegaal AC. 2015. SUMO in the DNA damage response. Oncotarget. 6(18):15734-15735.

Ingraham CR, Kinoshita A, Kondo S, Yang B, Sajan S, Trout KJ, Malik MI, Dunnwald M, Goudy SL, Lovett M, et al. 2006. Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). Nat Genet. 38(11):1335-1340.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 4(7):1073-1081.

Leslie EJ, Liu H, Carlson JC, Shaffer. JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, et al. 2016. Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies an Etiologic Missense Variant in GRHL3. The American Journal of Human Genetics 98(4): 744–754.

Leslie EJ, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, McHenry T, et al. 2016. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and19q13.Hum Mol Genet. 25(13):2862-2872.

Leslie E.J, Koboldt DC, Kang CJ, Ma L, Hecht JT, Wehby GL, Christensen K, Czeizel AE, Deleyiannis FW, Fulton RS, et al. 2016. IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. Clin Genet. 90(1):28-34.

Leslie EJ, Carlson JC, Shaffer JR, Butali A, Buxo CJ, Castilla EE, Christensen K, Deleyiannis FW, Leigh Field L, Hecht JT, et al. 2017. Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. Hum Genet. 136 (3):275-286.

Little J, Cardy A, Arslan MT, Gilmour M, Mossey PA. 2004. United Kingdom-based case-control study. Cleft Palate Craniofac J. 41(4):381-386.

Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, Becker J, Herberz R, Alchawa T, Nasser E, et. al. 2012. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet. 44(9):968-971.

Mangold E, Böhmer AC, Ishorst N, Hoebel AK, Gültepe P, Schuenke H, Klamt J, Hofmann A, Gölz L, Raff R, et al. 2016. Sequencing the GRHL3 Coding Region Reveals Rare Truncating Mutations and a Common Susceptibility Variant for Nonsyndromic Cleft Palate. Am J Hum Genet. 98(4):755-762.

Marazita ML, Field L.L, Cooper ME, Tobias R, Maher BS, Peanchitlertkajorn S, Liu YE. 2002. Nonsyndromic cleft lip with or without cleft palate in China: assessment of candidate regions. Cleft Palate Craniofac J. 39(2):149–156.

Marçano AC, Doudney K, Braybrook C, Squires R, Patton MA, Lees MM, Richieri-Costa A, Lidral AC, Murray JC, Moore GE, et al. 2004. TBX22 mutations are a frequent cause of cleft palate. J Med Genet. 41(1):68-74.

Mossey PA., Little J. 2002. Epidemiology of oral clefts: an international perspective. In Cleft Lip and Palate: From Origin to Treatment, D.F.Wyszynski, ed. (Oxford: Oxford University Press), pp. 127–144.

Mossey PA., Little J., Munger RG., Dixon MJ., Shaw WC. 2009. Cleft lip and palate. Lancet. 374(9703):1773–1785.

Peyrard-Janvid M., Leslie EJ., Kousa YA., Smith TL., Dunnwald M, Magnusson M., Lentz B.A., Unneberg P., Fransson I., Koillinen H.K., et al. 2014. Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am. J. Hum. Genet. 94(1):23–32.

Rauch GJ, Lyons DA, Middendorf I, Friedlander B, Arana N, Reyes T, Talbot WS. 2003. Zfin direct submission.

Richardson RJ, Dixon J, Jiang R, Dixon MJ. 2009. Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. Hum Mol Genet. 18(14):2632-2642.

Sabel JL., d'Alencon C, O'Brien EK, Van Otterloo E, Lutz K, Cuykendall TN., Schutte BC, Houston DW, Cornell RA. 2009. Maternal interferon regulatory factor 6 is required for the differentiation of primary superficial epithelia in Danio and Xenopus embryos. Dev. Biol., 325(1):249–262.

Sun Y, Huang Y, Yin A, Pan Y, Wang Y, Wang C, Du Y, Wang M, Lan F, Hu Z, et al. 2015. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun. 6: 6414.

Ting SB, Caddy J, Wilanowski T, Auden A, Cunningham JM, Elias PM, Holleran WM, Jane SM. 2005. The epidermis of grhl3-null mice displays altered lipid processing and cellular hyperproliferation. Organogenesis. 2(2):33-35.

Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G .2010. Protein structure analysis of variants causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics.11:548.

Waitzman NJ, Romano PS, Scheffler RM. 1994. Estimates of the economic costs of birth defects. Inquiry. 31(2):188–205.

Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, Meng L, Wang W, Song Y, Cheng Y, et al. 2017. Genome-wide analyses of non-syndromic cleft lip with palate identify fourteen novel loci and genetic heterogeneity. Nat Commun. 8:14364.

Table 1 Sequence variations for GRHL3 observed in CPO individuals from sub-Saharan Africa

HGVS	HGVp	Type	Ghana	Nigeria	1KG	EVS	ExAC	p	S	PS
c.332delC	p.Gly158Alafster55	Frameshift	0	1	0	0	0			
c.497C>A	p.Pro166His	Missense	1	0	0	0	0	PD	T	-0.64
c.1229A>G	p.Asp410Gly	missense/splice site	0	1	0	0	0	PD	D	-6.7
c.1282A>C	p.Ser428Arg	Missense	0	1	0	0	0	B	T	-1.46
c.1677C>A	p.Tyr559Ter	Stop-gain	1	0	0	0	0			

1Kg= 1000 Genomes, EVS= Exome Variant Server, ExAC= Exome Aggregate Consortium P=Polyphen, S=SIFT, PS= Provean Score, B= Benign, T= Tolerated, PD= probably damaging, D= Deleterious. c. refers to coding sequence position within the GRHL3 transcript NM_198173, amino acid substitutions refer to GRHL3 isoform 2, NP_937816

Table 2: Segregation analyses of novel mutations in case families

HGVS	HGVp	Individual	Genotypes
c.497C>A	p.Pro166His	case child	CA
c.497C>A	p.Pro166His	mother	CA
c.497C>A	p.Pro166His	grandmother	CA
c.1677C>A	p.Tyr559Ter	case child	CA
c.1677C>A	p.Tyr559Ter	Father	CA
c.1677C>A	p.Tyr559Ter	mother	CC
c.1229A>G	p.Asp410Gly	case child	AG
c.1229A>G	p.Asp410Gly	mother	AG
c.1282A>C	p.Ser428Arg	case child	AC
c.1282A>C	p.Ser428Arg	mother	AA
c.470delC	p.Gly158Alafster55	case child	C-
c.470delC	p.Gly158Alafster55	mother	C-

Figure 1. Rare variants in GRHL3 disrupt development of zebrafish EVL upon overexpression in zebrafish embryos. A) Schematic of the GRHL3 gene and the position of mutations identified in the present (red arrows) and previous studies (black arrows) (not to scale). All amino acid positions refer to GRHL3, variant 2 (NP_937816). The green arrow represents the missense mutation p.Thr454Met that is a common variant associated with non-syndromic CPO reported by Leslie et al., 2016 and Mangold et al., 2016. The yellow bar represents the DNA binding domain where the sumoylation and CK1 phosphorylation motifs are harbored in exon 4. The purple bar represents the dimerization region while the green bar represents the transcription activation domain. Blue boxes represent exons. **B-H**) Animal pole views of embryos fixed at 7 hpf and processed to reveal *krt4* expression. **B**) Embryos injected with *lacZ* mRNA, a negative control, have relatively even expression of *krt4* in all enveloping layer (EVL) cells (in three biological replicates of at least 40 injected embryos each, all had this appearance. None exhibited expression of *krt4* in deep blastomeres). **C**) Embryos injected with mRNA encoding a reference variant of human GRHL3 (AAH36890.1) develop with intense foci of *krt4* expression in deep blastomeres, recognizable in whole mount preparations as patches of dark staining (e.g., red asterisk) (in three separate experiments of at least 40 injected embryos each, all exhibited patches of ectopic *krt4* expression). **D, F**) Embryos injected with mRNA encoding the reference variant above altered to have the patient variants **D**) p.Gly158Alafster55 or **F**) p.Asp410Gly lacked ectopic patches of *krt4* expression, consistent with these being hypomorphic variants. Although *krt4* expression intensity appears lower than normal in this image, this was not consistently observed (in three biological replicates, a total of 99 total embryos injected, none exhibited patches of ectopic *krt4* expression. 10 or fewer non-contiguous deep cells expressing *krt4* in deep cells were observed in fewer than half of injected embryos). **E,H**) Rare embryos injected with **E**) p.Pro166His and **H**) p.Tyr559Ter variants exhibited patches of ectopic *krt4* expression (stars), indicating these variants retained activity

but at lower than wild-type levels (in three experiments, 4 of 88 and 5 of 91 injected embryos, respectively, exhibited patches of ectopic *krt4* expression as shown). **G**) About one third of embryos injected with p.Ser428Arg variant exhibited gaps in the expression of *krt4* (in three experiments, 23 of 64 injected embryos showed such gaps). Fractions: numerator is the number of embryos resembling the one in the image; denominator is the number of embryos injected. The residue in each case resembles un-injected embryos.

Figure 2. Injection of human *GRHL3* RNA partially restored *krt4* in the embryos injected with *grhl1* and *grhl3* MOs. **A-E**) Animal pole views of embryos fixed with the injected reagents at the one-cell stage, fixed at 6 hpf and processed to reveal *krt4* expression by in-situ hybridization. In embryos injected with **A**) control MO, **B**) *grhl1* MO, or **C**) *grhl3* MO, *krt4* expression is contiguous, whereas in embryos injected with **D**) both *grhl1* MO and *grhl3* MO *krt4* expression is highly reduced. **E**) In embryos injected with *grhl1* MO and *grhl3* MO and also with human *GHRL3* mRNA, there was mosaic rescue of *krt4* expression in the EVL, in addition to ectopic *krt4* expression in the deep layer cells. Fractions: numerator is number of embryos resembling the injected one (in E, includes ectopic *krt4*-expressing cells), denominator is the total number of embryo injected. In **D**, the residual embryos resembled control MO-injected ones. In **E**, the residual embryos resembled those injected *grhl1* MO and *grhl3* MO pictured in **D**. **F** qRT-PCR analysis of *krt4* mRNA levels in embryos injected with the indicated MO and indicated mRNA. **, significant (p<0.01) in comparisons to the "*grhl1* MO+*grhl3* MO+lacZ" group using a one-way ANOVA. Triangles, significant (p<0.01) in comparisons to "*grhl1 MO+ grhl3 MO+GRHL3 group*".

Figure 3. In vitro reporter analysis of transactivation activity of rare variants in GRHL3. A) Bar chart representing firefly luciferase activity, normalized to Renilla luciferase activity, in 293FT cells co-transfected with the indicated (0.5 µg) GRHL3 expression constructs, or pCS2+ plasmid, together with 0.5 µg of a GRHL3-sensitive firefly reporter plasmid and 0.05 ug a constitutive Renilla reporter plasmid as a transfection control. **B**) Bar chart representing firefly luciferase activity, normalized to Renilla luciferase activity, in 293FT cells co-transfected with 0.5 µg indicated GRHL3 expression constructs (or CS2+) along with 0.5 µg wildtype GRHL3 expression constructs, the GRHL3-sensitive firefly reporter plasmid, and the Renilla reporter plasmids. p.Arg345Cys is a rare GRHL3 variant found in Van der Woude syndrome (referred to as p.Arg391Cys in Peyrard-Janvid et al., 2014), and p.Thr408Met is the CP-associated *GRHL3* variants reported previously (referred to as pThr454Met in Leslie et al., 2016, Mangold et al., 2016). These two variants serve as positive controls for dominant negative activity for the current study. The data represent results from three separate experiments, and error bars represent standard error. Asterisk indicates $p < 0.05$.